Towards an Inclusive Approach to Evaluation of Teaching

Valentine Olusegun Matthews[1]

[1] Lancaster University, UK

Correspondence: Valentine Olusegun Matthews, Lancaster University, UK.

Received: August 7, 2025 Accepted: September 6, 2025 Online Published: September 8, 2025

doi:10.5430/ijhe.v14n5p1 URL: https://doi.org/10.5430/ijhe.v14n5p1

Abstract

This research seeks to further knowledge on student and faculty evaluation of teaching within a small Asian university, with the aim of exploring the relationship between students and faculty perceptions of teaching. It seeks to establish faculty perception of fairer mechanisms or collective mechanisms for evaluating teaching, learning, and curriculum course materials towards evaluating the faculty. This contrasts with common use of student evaluation of teaching as a single approach. Faculty perceptions of student evaluations of teaching have been well researched in several contexts but less so in Middle Eastern universities with highly transient faculty drawn from different countries across the world.

Mixed methods, secondary and primary data of student and faculty evaluations of teaching are studied. Course materials and instruction ratings are found to be moderately correlated. Faculty are satisfied with the use of the student evaluation and faculty self-evaluation of teaching. However, outcomes of faculty self-evaluation and student evaluation of teaching are moderately positively correlated. Outcomes of student and faculty evaluations of teaching are found to be clustered at the upper end of the Likert scale. The implication is that there are small differences between scores from student evaluation and faculty self-evaluation of teaching. However, findings from the qualitative comments provide more valuable policy arguments. The findings indicate that half of the faculty agree or strongly agree with the use of SETs for customer satisfaction or as a control tool. However, a model that combines multiple evaluation tools is suggested by faculty for a holistic evaluation.

Keywords: evaluation, teaching, self-evaluation, peer evaluation, faculty evaluation

1. Introduction

1.1 Background

Student Evaluation of Teaching (SET) surveys are commonly employed for evaluating the quality of teaching and learning (Gustad, 1961; McKeachie, 1997;) in a Higher Education Institution (HEI). Its scope of use is dependent on the quality context (Hou et al., 2022). Its widespread adoption is attributable to its ease of implementation and interpretation (Penny, 2003). Theoretically, outcomes of SETs have had significant implications for quality of teaching (Johnson, 2000). SETs are also used in performance evaluation of faculty and for querying student voice for teaching, learning and decision making (Lyde et al., 2016). Tools such as Faculty Self-Evaluation of Teaching (FSET), Faculty Peer Evaluations and Independent Consultant Evaluations (ICE) (Borgen and Davis, 1974). These are often implemented to complement SETs (Benton and Young, 2018).

The SET content is not universal and is specific to individual HEI. To extract pertinent information that informs improvements, implementation and interpretation of SET is done at course, program or at whole HEI levels (Arbula Blecich & Zaninović, 2019; Schultz and Latif, 2006). This suggests that conducting the SET at the micro-social domain or at the level of the faculty extracts student perception of the course. However, for the purposes of programs and institutional monitoring, the analysis is at the macro-level. Evaluating teaching at institutional level produces overall HEI information but suffers from significant loss of information at the policy level. The outcomes of the SETs are most pertinently deployed at the level of the faculty for improvement of teaching, faculty evaluation and other student concerns.

Relationships between outcomes of SET, FSET, Peer Evaluation and ICE could inform the usage of these tools for decisions on faculty performance. The key challenge is to explore relationships between FSET and SET, and explore faculty preference for SET, FSET, peer evaluation, ICE or a combination of these for effective faculty evaluation.

1.2 Purpose of the Study

It is essential to explore the likelihood that the FSET and SET outcomes are similar. In such scenarios where these are similar, the correlation would further support the justifications for adopting the SET. It would indicate that the student perceptions of the teaching captures similar results as the perceptions of the faculty. It would also indicate that combining both tools in evaluation would be redundant. Thus, it is necessary to explore faculty positions on the ICE, SET, FSET and peer evaluation or a combination of these as a more holistic approach for evaluating the faculty. The research questions include:

(1) What does the outcome of the SET reveal about the effectiveness of teaching?

(2) Is there a relationship between the evaluation of the instruction and the course?

(3) How do faculty perceive the SET in evaluation?

(4) How do they perceive other assessment tools?

(5) Challenges to the use of the SET for faculty evaluation?

(6) Are the FSET and SET outcomes related?

(7) Alternatives to use of SET for faculty evaluation

2. Literature Review

2.1 Effective Teaching

Effective teaching has been subjectively defined and is said to be known by observers when seen (Benton and Cashin, 2012; Cashin, 2003; Miron & Mevorach, 2014). Effective teaching is sometimes seen as the propensity to improve student achievement. This approach perceives the existence of a cause-and-effect relationship between teaching and student performance. Effective and excellent teachers may possess attributes such as dynamic teaching styles, accessible personal traits and clear teaching techniques (Alhijia, 2017; Pan et al., 2009). They could also demonstrate good teacher-student relationships, engage and have real-world experience (Keeley et al., 2015; Pan et al., 2009). Institutional accountability requires HEIs to implement adequate measures for assuring stakeholders that the quality of teaching is being improved (Kember et al., 2002).

2.2 Student Evaluation of Teaching (SET)

SETs are the foremost tools for evaluating teaching or teaching effectiveness (Gustad, 1961; McKeachie, 1997; Spooren et al., 2013). In a Turkish study, SET data collected from students was seen as valid tool for evaluating faculty performance (ÜSTÜNLÜOĞLU & Seda, 2012). Faculty development (Park et al., 2020; Theall, 2017) are informed by the outcomes which also serve as evidence of institutional teaching effectiveness (Denson et al., 2010). The SET is sometimes used as the sole means for faculty evaluation. HEIs also invest significant efforts on SETs to address concerns of student satisfaction with teaching and programs (Richardson, 2005). SETs are also implemented in courses as a quality assurance tool to survey student satisfaction with the instruction in courses (Johnson, 2000). This is likely with the view that effective teaching includes effective instruction and course materials.

SET outcomes, student characteristics, grading leniency, instruction and learning have been found to be highly correlated (Arbula Blecich & Zaninović, 2019; Spooren 2010). On the contrary, some studies found SET outcomes to be independent of student gender (Esarey and Valdes, 2020), grades and field of study (Spooren and Van Loon, 2012) with the bias being against female teachers (Boring, et al., 2016). Basow and Silberg (1987) found that independent of student gender, female faculty are rated lower than males. Course discipline, sexual orientation and race of the faculty are also seen as biasing factors of SET outcomes (Spooren et al., 2013). Such biases and other factors not tested by the SET could likely skew the SET outcomes. There is no universally tested and agreed SET, as their contents of the questions are often context specific. SET outcomes have been found to be biased by faculty personality (Shah et al., 2020). Chen & Hoshower (2003) found students who positively appraised their professors generally viewed the course evaluation system more positively. It is important to establish if this relationship exists in a less liberal education context in the Middle East.

SET informs decisions on faculty salary increases and retention (Benton & Young, 2018; Kulik, 2001). Faculty have also disagreed with outcomes of SETs for promotion or salary increases (Iyamu & Aduwa, 2005). With so many biases, the outcomes of the SETs are often disputed by faculty. SET outcomes have sometimes shown that faculty with best reviews were not necessarily the most effective (Surratt & Desselle, 2007). Students' lack of training in implementing course materials is also seen by faculty as an invalidating factor of the SET outcomes (Mohammad, 2001). However, SET queries wider experiences than the curriculum.

SETs are commonly implemented by Middle East and Asian HEIs as instruments for measuring teaching effectiveness. This is reflected in Jordan, Omani and Chinese contexts (Alsmadi, 2005; Yin et al., 2016). Faculty gender and class size were identified in an Omani study as weak predictors of SET outcomes. Faculty have not unanimously received the SET (Iyam & Aduwa-Oglebaen, 2005; Mwachingambi & Wadesango, 2011) while other publications in Africa and Asia show that faculty accept the SET only for formative purposes (Idaka et al., 2006; Inko-Tariah, 2013). The initial step is to explore the outcomes of SET within a small university context in the Middle East.

2.3 Faculty Perceptions of the SET

SETs are often criticised by faculty over their insufficiency in identifying effective teaching (Boring et al., 2016; Hornstein, 2020). Its outcomes are used to make decisions on promotion, salary increases and retention of faculty, hence the crucial need to capture their perception of the tool. Faculty also raise concerns over its validity, reliability of outcomes, (Boring et al., 2016, Hornstein, 2020 and Spooren et al, 2013) students, faculty and course characteristics (Spencer & Schmelkin, 2002; Wachtel, 1998). These challenges of reliability, validity and concerns with SETs have not impeded their adoption by HEIs. Faculty perceptions of SETs, their purpose, validity, usefulness, and consequences remain mixed (Wachtel, 1998). The faculty perceptions of the SET, its challenges and its sufficiency in evaluating faculty performance are desired. The evaluations, development, salary and promotion of the faculty depend largely on the outcomes of the SET. It is thus necessary to understand their perceptions of the tool and other evaluative tools such as FSET, peer evaluations, and ICE. A challenge is to uncover faculty perception of the SET in a highly transient context with very diverse collective of faculties from university cultures across the world.

2.4 FSET and other Evaluative Tools

FSET describes faculty self-perception of quality of their own teaching (Lyde et al., 2016). There are indications that faculty are satisfied with the use of SETs as a tool for formative feedback on the instruction and instructor (Balam & Shannon, 2010; Kulik, 2001; Spooren, 2013). This suggests that there is a relationship between SET outcomes and faculty perception of their teaching. Where there is agreement between the SET and FSET outcomes, combining both tools in evaluation of faculty would be redundant. Feldman (1989) found correlation between SET and FSET to be low while medium correlation below 0.5 was found by other studies (Beran and Violato, 2005; Marsh & Dunkin, 1992). Although the sample size is small, the low correlations are indicative that the faculty and student views of SET outcomes do not strongly align. The effectiveness of FSET is likely a challenge due to its subjectivity which again raises questions of its use for formative evaluations, as an accountability tool (MacBeath, 2003) and for faculty performance evaluation (Atwater et al., 2002; Lyde et al., 2016; Moreno et al., 2021). There are risks associated with the investment of trust in faculty to objectively self-assess. HEI's tend to adopt a combination of feedback sources to inform performance evaluation (Entrekin & Chung, 2001). Use of these combinations of evaluative tools may be challenged where strong correlations exist between some of these tools.

Faculty have been found to use peer evaluations for formative developmental purposes (Berk, 2005). The criticism is that the reports may be skewed by peers with insufficient experience of quality teaching (Chism, 1999). FSET involves the use of structured tools, response to structured questions or reflective statements by faculty on their pedagogical practices. Such reflective tools support the faculty's professional growth through self-awareness for the purpose of alignment of practice with policy goals. FSET can reveal insights into teaching philosophy and strategies, complementing other tools like the SETs. Its subjectivity is the main limitation. It is open to faculty bias or lack of critical self-assessment skills (Marsh and Roche, 1997).

ICE is debatably more objective than the SETs for the evaluation of teaching. Consultants would typically be experts in pedagogy, in assessing instructional materials, classroom methods, and learner engagement using structured observations and rubrics. Hence, this would suggest the existence of a recognized teaching evaluation framework would be applied. Where publications such as Esarey and Valdes (2020) noted gender-based biases in SETs, ICEs are evidence-based and standardized. ICEs are based on a "fairness outlook" for tenure, salary and promotion decisions. However, this suggests that the faculty needs to be trained to effectively apply these prescribed teaching standards. Berk (2018) noted that limited expertise in pedagogy and other aspects of teaching can lead to superficial assessments. ICE may also suffer from inadequate knowledge of the local context, especially if as in the case of the UAE, the consultants are commonly foreign.

3. Methodology

The methodology is designed to address the research objectives by exploring secondary data of the SET outcomes within courses. FSET is also run to capture faculty perceptions of their teaching using the same SET questions. Faculty perceptions of other evaluation tools are queried using surveys. The section is presented in the order below:

3.1 Research Design

Secondary data is retrieved from the HEI while primary faculty FSET and perceptions survey are retrieved from faculty. The HEI's academic units are all within the field of business hence a common pedagogy is defined in policy. The primary and secondary data are subjective constructs of the social reality of the faculty and students in this context (Flick, 2009; Shah et al., 2020). The survey is completed by a purposive sample of 11 of the HEI's 20 faculty members who volunteered to participate in the research. The data are either Likert ratings or qualitative data. The Likert data are analysed using probabilities and averages while the qualitative data are collated summarized and presented.

3.2 Method

Mixed method is often employed in researching SETs (Chan et al., 2014; Chikazinga, 2018; Iyamu and Iyamu-Oglebaen, 2005). It is similarly adopted in this research because it is largely more suitable for extracting meso- and macro-level information (Chen et al., 2014; Lyde, 2016). While the quantitative approach provides collective insight into specific closed-ended questions, the qualitative approach provides clearer insights into performance of individual faculty or in the case of the questionnaire, further insight into the closed ended questions. Macro-level study requires the use of larger data sets for making inferences (Chen et al., 2014; Ginns et al., 2007; Sulong, 2014) however, this study is a meso-level study and employs combined primary and secondary mixed data.

3.3 Research Context

The case HEI is situated in the Middle East region. The region is actively improving its higher education sector in response to global competitions for skills development. The HEI is a small Business School, established about 15 years prior and delivers graduate level business and leadership related academic programs. It has similar characteristics to the research context of Basow and Silberg (1987) who researched the SET in a 4-year liberal arts college in the US. SETs are commonly implemented by HEIs in the USA (Osburne & Purkey, 1995), Asia and Europe (Alsmadi, 2005; Chen & Hoshower, 2003), and Nigeria (Iyam & Aduwa-Oglebaen, 2005). These publications focused on large higher education institutions with conglomerated responses from multiple fields and faculties. This study focuses specifically on graduate students' SET outcomes and graduate level faculty within the business field. The case HEI is characterized by having at most 20 academic faculty members at any time. The academic programs are graduate degrees with a length of 1.5 years and the maximum number of students in an academic year is 400. The HEI is also strongly characterized by high faculty turnover. The Faculty have an average of 8 years' graduate teaching and research experience in the higher education sector.

3.4 Participants

Students and faculty are good sources of informed data on the quality of teaching (Cashin, 1989; Chism, 1999). Secondary SET data from the HEI are consulted to capture student opinion of the instruction and course materials of 11 out of the 20 faculty members of the Business School. FSET with the same questions as the SET are implemented in the same faculties. The questions include Likert-style rating of closed ended questions and qualitative response to an open-ended question seeking additional reflector-knowledge.

3.5 Instrumentation

Table 1. Contents of the case HEI's SET

Course Material Descriptors	
A1. The course has clear objectives.	B1. The instructor was knowledgeable about the course topics.
A2. The order of the course contents made sense.	B2. The instructor was enthusiastic about the course topics.
A3. The quality of the course materials (slides, readings, exercises, case studies etc.) for this course were of a high standard.	B3. The instructor was responsive to student questions.
A4. The course components (e.g., sessions, tutorial sessions, coursework, etc.) enhanced my understanding of the topic.	B4. Overall, the instructor was effective.
A5. Overall, the course was effective in delivering its Learning Outcomes.	B5. The instructor helped me to understand the course content clearly.

Table 1 shows the questions from the HEI's SET survey. For the analyses, aspects A1 … A5 represents all questions from section A (Course materials). These questions are designed to retrieve student perceptions of the course. Aspects B1 … B5 represent all questions from section B (instruction descriptors). These questions are designed to retrieve student perceptions of the instructor's teaching quality. In addition to these questions, the written-in comment question below is added to the survey: "What do you think could be improved in the instruction?". The SET questions are also adapted to create the FSET to capture faculty perceptions.

The faculty perception questionnaire has 32 questions clustered into 7 categories: perception of the implemented SET, perceptions of alternatives to SETs (Henderson et al., 2014; Douglas and Douglas, 2006), ICE, FSET, factors challenging use of SET for performance evaluation, disadvantage of the implemented SET, and qualitative comments. The questionnaire was adapted from Deaker et al. (2010) with a mix of Likert scale, Yes/No, and "open-ended comments" questions. Likert-type scales are ordinal scales of series of qualitative statements that respondents use to indicate and convey their opinion or perceptions of evaluative questions (Vogt, 1999). The Likert scale adopted in this study range from strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4) and strongly agree (5), piloted by 6 faculty members. Cronbach alpha from this pilot is 0.714. Cronbach Alpha reliability coefficient close to 1 validates the faculty perception of SET questionnaire (Harun et al., 2011; Iyam and Aduwa-Oglebaen, 2005; Yin et al., 2016).

The SET and the FSET questions reflect the aspects most relevant to the context, educational vision, policy, and in accordance with its perception of teacher quality (Penny, 2003). The STE/FSET has three sections: Part A (course), Part B (instruction) and Part C (comments). Sections A and B together have a Cronbach Alpha of 0.926 and contain 5 questions each with a 5-point Likert scale response. Section C as aptly titled is a written-in comment section. The SET is an anonymized survey implemented on 11 courses of the 11 faculty by the institution during the course and prior to the summative assessment. The FSET is implemented over the same 11 faculty on the same 11 courses.

3.6 Data Collection

The faculty are purposively sampled with 11 of 20 faculty members. 3 of the 4 females are included in the study. 4 Assistant professors, 2 full professors and 5 associate professors. 3-4 faculty from each of the 4 academic programs participated in the study. Purposive sampling approach was adopted because the sampling population of faculty is 20. However, participation in the study by 11 members was voluntary. Table 1 describes the characteristics of the participants from all 4 academic programs. The faculty mean experience at the HEI is 3.11 years and their mean experience in HEIs generally is 13.28 years. This implies that the faculty possess significant experience in higher education.

3.7 Data Analysis

In answering the research question "what does the outcome of the SET reveal about the effectiveness of teaching?" and "is there a relationship between the evaluation of the instruction and the course?", secondary data captured by the HEI at the end of course delivery of 11 courses are analysed using descriptive statistics. Means instructor and course material scores are found for each course. A linear regression is implemented between the ratings of the

course materials and the course instructions. Students' written-in comments in the SETs are coded, categorized and themes are created. The themes hold collective meanings and are 'outcome of coding, [and] categorization' (Saldaña, 2013, p. 14).

To address the question on "how the faculty perceive the SETs in faculty evaluation", "perception of other assessment tools", and "challenges to the use of the SET for faculty evaluation", the data collected from the questionnaires analysed by calculating the probability that the ratings P(X) between 1 and 5. It provides visuals of the highest ratings. To determine faculty order of preference of the SET, FSET, Peer Evaluation or ICE, the ranked scores are added to determine the collective faculty scores.

To address the question of existence of a relationship between the FSET for each faculty and their average course SET scores are determined. Average faculty rankings of the SET, FSET, ICE, and peer reviews for faculty evaluation are presented using a bar chart. Faculty perception of the use of a combination of these tools is also presented to determine their position on a combination of these tools for faculty evaluation.

4. Results

4.1 SET in Evaluating Teaching Effectiveness

The average class size is 24 and 296 students participated in these courses. The overall student participation rate in SET implemented on the randomly selected courses is 68%. The mean student ratings of the effectiveness of the course materials and instruction are between 4.45 (0.25) and 4.54 (0.37) respectively.

Table 2. Summary of SET survey outcome for 11 courses taught by the surveyed faculty

Faculty	*Participation rate	Course Mean	Instruction means
X1	0.92	4.65 (0.18)	4.93 (0.03)
X2	0.63	4.49 (0.05)	4.61 (0.06)
X3	0.69	4.39 (0.09)	4.29 (0.05)
X4	0.63	4.47 (0.10)	4.50 (0.07)
X5	0.75	4.58 (0.13)	4.87 (0.04)
X6	0.51	4.36 (0.07)	4.35 (0.02)
X7	0.81	4.65 (0.06)	4.78 (0.05)
X8	0.78	4.32 (0.02)	4.58 (0.02)
X9	0.67	4.55 (0.13)	4.65 (0.14)
X10	0.75	3.78 (0.18)	3.62 (0.09)
X11	0.73	4.68 (0.07)	4.77 (0.05)
Mean	0.68	4.45 (0.25)	4.54 (0.37)

*Participation rate = No. of responses/ to class size

Table 2 presents the outcomes from the analysis of the secondary SETs implemented midway through delivery of 11 courses by faculty members. The results suggest that on average, students agree or strongly agree that the faculty's course materials and instruction are effective. Outcomes of faculty X10 are outliers due to its low mean rating of 3.78 ± 0.18 for the course materials and 3.62 ± 0.90 for the instruction. The ratings suggest that faculty X10's course materials and instruction are viewed by students as less than effective. Comments on the SET for faculty X10 show that the students feel the teaching materials are too many and too technical. The students indicated that the teaching and materials need to be further simplified.

Table 3. SET course materials and instruction outcomes

(A) Course Material		(B) Instruction	
Descriptor	Mean (SD)	Descriptor	Mean (SD)
A1	4.54 (0.27)	B1	4.56 (0.36)
A2	4.47 (0.30)	B2	4.54 (0.33)
A3	4.41 (0.22)	B3	4.57 (0.40)
A4	4.40 (0.30)	B4	4.53 (0.37)
A5	4.41 (0.27)	B5	4.51 (0.39)
Mean	4.50 (0.25)	Mean	4.54 (0.37)

Table 3 presents the outcomes of the analysis of the HEI's aspects of course materials and instruction. The mean ratings of the course evaluation question A1 … A5 are generally lower than the scores of the instruction evaluation questions (B1, … B5). At a minimum, the students generally view the course materials and the instructions as effective.

4.2 Relationships between SET Course and Instruction Ratings

Results of the SET reveals the outcomes of "section A" which reveals the student's evaluation of the faculty's course materials and the outcomes of "section B" or the instruction. The outcomes of the analysis presented here explored the relationships between students' rating of the course materials and the instructions. The outcomes of students' ratings of the instructional materials and the instruction are presented in Table 3 however, the relationship between these sets of ratings are presented in Table 4.

Table 4. Relationship between course and instruction ratings

	Coefficients	Standard Error	t-Stat	P-value
Intercept	1.464	0.328	4.466	0.00156
Instruction	0.657	0.072	9.126	0.00001
Note: $r = 0.95$, $\alpha = 0.05$, Depending Variable = Course Materials				

Table 4 provides the outcomes of the implemented regression analysis between the means of student rating of the course materials and the instruction. This seeks to establish the degree of relationship between the students' ratings of the course materials and the instruction. With $r = 0.95$, $\alpha = 0.05$ and the p-value 0.00001, the findings show that the ratings of the course materials and instruction are strongly correlated. The finding agrees with Spooren (2010) that the course ratings and instruction ratings are similar however, the findings deviate regarding the strength of the relationship. It could be argued that the SET as implemented, assesses students' opinions within the "happiest period" of the course because the final course grades which would likely impact their opinions has not been received. Students would likely blame poor performance on the instruction.

4.3 Faculty Perceptions of the SET

Outcomes of the faculties' perceptions of the SET retrieved from the implemented questionnaire are described in this section.

Table 5. Faculty perceptions of the HEI's implemented SET survey

Cluster	Questions	*P (X)					Mean
		P(1)	P(2)	P(3)	P(4)	P(5)	
SET and the outcomes in the HEI context	(1) The outcomes of the SET accurately describe my teaching.	0.09	0.09	0	0.64	0.18	3.73 (1.14)
	(2) The questions in section B (instruction) of the SET are sufficient for evaluating the teaching skills of a faculty.	0.09	0.18	0	0.64	0.09	3.45 (1.16)

*P(X) = Probability of rating the questions on a Likert scale with scores X = 1, 2, 3, 4, 5.

Table 5 provides the outcomes of the faculty perceptions of the contextual SET survey. Faculty perception is that faculty tend to "somewhat agree" that the SET outcomes reflect their teaching thus suggesting that the SET and FSET outcomes are somewhat correlated. The probability P (>4) = 64 shows that 82% of the faculty somewhat agree that the outcomes of the SET accurately reflect their teaching.

The qualitative comment section provides faculty unrestricted opportunities to express their opinions about the implemented SET in the HEI. By using qualitative classification method (Bailey, 1994) after coding, the comments are classified into collections of meanings. Faculty recommendations for improving the HEI's SET include the separation of the course materials and instructor surveys to reduce effect of one on the other and to use more benchmarked questions. This further emphasizes the high correlation found between the course materials and instruction correlation. Mean rating of 3.45 ± 1.16 in question (2) suggests that the faculty neither agrees nor disagrees with the assertion that the questions in the SET are sufficient for evaluating the teaching skill of the faculty.

The SET question solicited student responses on how the instruction and materials could be improved. Five categories were generated with seventeen codes. The categories include delivery and assessment, poor teaching materials, preparation of lesson resources, teaching styles and lesson structure. Although the students agree or strongly agree with the effectiveness of over 90% of the instruction and courses, the qualitative comments provide further insights into changes proposed for further improvements of the courses and instructions.

Table 6. Outcomes of thematic analysis of the written-in comments in the SET

Themes	Codes
Delivery and assessment time	• Extend assessment submission date
	• Increase delivery time
	• Reduction in the length of lessons (x3)
	• Assessments time constraint
Poor teaching material	• Abstract exercises
	• More details in materials
	• Simplify the materials
	• Improve teaching materials
Preparation of lesson resources	• Prepare microphones before lessons
	• Provide software before sessions
Teaching styles	• Increase use of video
	• Make lessons more interactive (X3)
Lesson structure	• Unsystematic lesson structure

Table 6 presents the outcomes of students' response to this research question of how the SET could be improved. The themes represent the outcomes of the analysis.

4.4 Faculty Perceptions of Alternative Tools

Outcomes of faculty perceptions of alternative tools for assessing the effectiveness and quality of teaching are provided in this section.

Table 7. Faculty perceptions of alternatives to the SET survey

Cluster	Questions	*P (X) P(1)	P(2)	P(3)	P(4)	P(5)	Mean
Faculty Perceptions of Alternatives to SET	(3) Peer evaluation could be acceptable using a standardized evaluation framework.	-	0.27	0.27	0.09	0.36	3.55 (1.23)
	(4) ICE is more acceptable for assessing teaching than SET.	0.46	0.18	0.36	-	-	1.91 (0.90)
	(5) FSET is more acceptable for assessing quality of teaching than SET.	0.09	0.27	0.18	0.36	0.09	3.09 (1.16)
	(6) Peer evaluation of teaching is more acceptable for assessing teaching than SET.	0.27	0.36	0.18	0.18	-	2.27 (1.05)

*P(X) = Probability of rating the questions on a Likert scale with scores X = 1, 2, 3, 4, 5.

The questions in Table 7 elicit faculty perceptions of implementing alternatives to SET survey. Question (3) explores faculty perception on the use of standardized teaching framework for evaluating faculty teaching quality. The mean of 3.35 ± 1.16, shows that 45% of the faculty strongly or somewhat agree with the use of a standardized framework for evaluating teaching over the SET. With a mean of 1.91 ± 0.90, the faculty disagree or are neutral to the use of ICE in the evaluation of effectiveness of teaching over the SET. A third of the faculty strongly disagree with the use of FSET over SET. The mean value of 3.09 ± 1.16. However, 45% of faculty agree or strongly agree with the use of the FSET over the SET. Against the SET, the ICE is the least preferred tool. This suggests that the faculty prefer the use of SETs over other evaluative tools.

4.5 Faculty Preferred Evaluative Tools

The ranked scores seen in figure 1 show that faculty ranked the SET above the other tools for the purpose of effectively evaluating teaching. The FSET and peer review scores are not too far back suggesting that faculty placed intrinsic value in these tools. The use of ICE is the least preferred by the faculty. Faculty perception is that FSET is viewed as more effective in teaching evaluation than peer evaluation but would accept peer evaluation if a standardized teaching evaluation framework is adopted. Findings in Table 7 suggest that ICE, FSET and peer evaluations are less preferred to SET. These findings further buttress the rankings seen in figure 1.

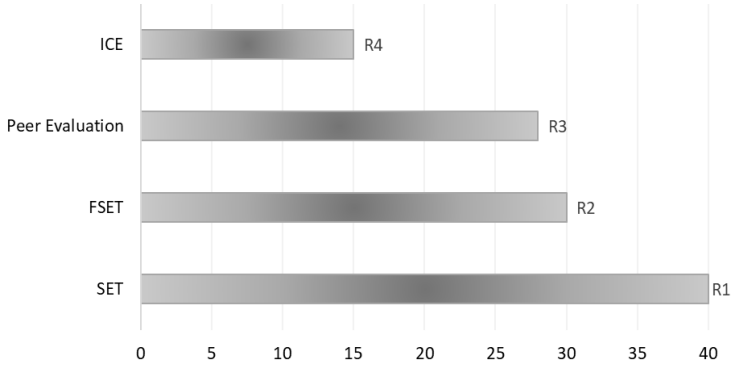

Figure 1. Faculty preference of SET, FSET, ICE and peer evaluations

Figure 1 presents the outcome of the order of preference of the SET, FSET, ICE and peer evaluations as ranked by the members of faculty. Adoption of this "diverse input mode" of the SET and any or a combination of FSET, peer evaluations and ICE could improve the robustness of the evaluation of teaching. Using only the SET and not the

other tools may not be adequately robust, depending on how closely the other tools are ranked relative to the SEN. 27% of the faculty did not view the SET as the highest ranked tool and 73% of the faculty view the FSET and peer evaluation as their sent ranked tool.

4.6 Challenges to the Use of SET for Faculty Evaluation

This section presents findings on the challenges impacting the use of SETs.

Table 8. Faculty perceptions of factors challenging the use of SET

Questions	*P (X)					Mean
	P(1)	P(2)	P(3)	P(4)	P(5)	
(7) Students are not sufficiently skilled to evaluate the teaching skills of a faculty.	0.18	0.46	-	0.27	0.09	2.64 (1.30)
(8) SET should have a legitimate place as a form of consumer control, and less as a faculty evaluation tool.	-	0.18	0.27	0.18	0.36	3.73 (1.14)
(9) Faculty charisma is more influential in SET than the subject of the lesson.	0.09	-	0.27	0.27	0.36	3.82 (1.19)
(10) The lecture content has little effect on the student ratings.	0.36	0.18	0.27	0.18	-	2.27 (1.14)
(11) I am in favor of SET if it is not used for performance appraisal of the faculty.	-	0.46	0.09	0.27	0.18	3.18 (1.19)
(12) There are important aspects of teaching which cannot be evaluated by simply rating statements on a Likert Scale.	-	0.09	0.27	0.27	0.36	3.91 (1.00)

*P(X) = Probability of rating the questions on a Likert scale with scores X = 1, 2, 3, 4, 5.

Table 8 shows faculty perceived factors challenging the use of SET for faculty performance evaluation. 64% of the faculty somewhat or strongly disagree that students are not sufficiently skilled to evaluate the teaching skills of the faculty. This conflicts with the findings of Mohammad (2001) who viewed students' lack of training on course materials as an invalidating factor of SET outcomes. As graduate students, the students would have significant background in the subject and numerous teaching approaches. Hence their scoring of the faculty is likely to be sufficiently informed. Students likely have curriculum and implementation skills to evaluate the effectiveness of the lessons. It is arguable that as customers and graduate students, they may be in the best position to know if the lessons meet their needs.

54% of faculty somewhat or strongly agree that the SET has a legitimate place as a consumer control tool and not as a performance evaluation tool. This suggests that although it is preferred to the other tools, faculty still hold some concerns with the SET. 63% of faculty believe SET outcomes are more influenced by faculty charisma than the subject of the lesson. Faculty charisma may go a long way in maintaining student attention but does not effectively address knowledge gained or the quality of the course materials. 63% of faculty somewhat or strongly agree that the lecture content is more influential in the SET ratings.

Outcomes of the categorized qualitative comments indicate that faculty believe that positive bias would be introduced by friendly peers or negatively introduced due to faculty feud or disagreements. This suggests that peer reviews are skewed by internal relationships. They also perceived that peer evaluations could complement other tools used to inform performance evaluation and should not be used as a stand-alone evaluative tool. If objectively implemented, peer reviews are also believed to inform decisions on professional development.

Table 9. Further challenges to SET implementation

Questions	*P (X) P(1)	P(2)	P(3)	P(4)	P(5)	Mean
(13) The SET negatively influences my decision in choosing a teaching approach.	0.18	0.46	0.09	0.27	-	2.45 (1.08)
(14) SET negatively influences how I relate to the students.	0.27	0.36	0.09	0.18	0.09	2.45 (1.30)
(15) SET outcomes negatively influence the design of my course materials.	0.46	0.36	-	0.18	-	1.91 (1.08)
(16) Allowing for anonymous student response encourages silly and amusing responses.	0.09	0.36	0.36	0.09	0.09	2.73 (1.05)
(17) Students' ratings on the SET can be unfairly skewed by ratings of "extreme students".	-	0.09	0.09	0.46	0.36	4.09 (0.90)
(18) Student assessment could negatively impact the SET.	-	0.27	0.27	0.18	0.27	3.45 (1.16)
(19) The student comments in the "comment section" were generally constructive.	0.09	-	0.18	0.55	0.18	3.73 (1.05)
(20) The students' comments were generally the most useful segment of the survey.	0.09	-	0.09	0.46	0.36	4.00 (1.13)
(21) Written-in adverse comments were the most useful responses for faculty improvements.	0.09	-	0.27	0.46	0.18	3.64 (1.07)

*P(X) = Probability of rating the questions on a Likert scale with scores X = 1, 2, 3, 4, 5

Findings seen in Table 9 indicate that faculty relationships with the students and the choice of teaching methodology are not negatively impacted by the SET. Faculty also are not affirmative that the outcomes of SETs influence the design of the courses. In other words, faculty do not steer the course materials to game the SET outcomes. The faculty also disagrees with the notion that anonymization of the SET opens the responses to silly and unhelpful comments. This suggests that the activity is taken very seriously by the students. 82% of the faculty members, however somewhat or strongly agree that extreme responses skew the SET ratings.

Ratings of question (18) indicate that 45% of faculty believe that summative assessment negatively impacts the outcomes of the SET however, 55% are either unsure or somewhat disagree that it has an impact. This suggests that the faculty are split on the impact of student assessments on SET scores. This view of the faculty is likely skewed by the fact that the HEIs conduct these SEN feedback well in advance of the assessments. With a mean rating of 3.73, at least 73% of the faculty somewhat or strongly agree that the students' comments are generally constructive. 82% somewhat or strongly agree that the comment section is the most useful section of the SET while only 64% view it as most useful for developing faculty.

4.7 FSET and SET Outcomes: do they Agree

Table 10. Regression of the FSET and SET outcomes

	Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
	B	Std. Error	Beta	t	p	Lower Bound	Upper Bound
(Constant)	-5.032	9.839		-0.511	0.644	-36.343	26.279
SET	2.105	2.166	0.489	0.972	0.403	-4.788	8.999

Note: r = 0.489, α = 0.05, Dependent Variable: FSET

Table 10 shows the outcomes of the regression analysis. With correlations coefficient (r) = 0.489 and p-value = 0.403, SET and FSET outcomes are moderately positively correlated however, this relationship is not significant at significance level (α) = 0.05, hence the model is not a good model and SET scores do not reflect the FSET ratings. This suggests that although faculty are satisfied with the SET outcomes, their ratings of the FSET only moderately reflect students' ratings of teaching.

5. Discussion

5.1 The SET and its Outcomes

The SET findings agree with Debroy et al., (2019) that students rating of faculty is commonly >4. These consistent set of readings suggest that the student ratings are biased or do not provide sufficient discrimination of good practice from poor practice. These consistent >4 SET ratings agree with the findings of Dolnicar & Grün (2009), Clayson (2009) and Onwuegbuzie et al. (2009) that extreme student ratings bias the SET outcomes due to tendencies of some students to rate at the extremes. The strong correlations between the mean scores of students' course materials and instruction ratings agree with Spooren & Mortelmans (2016), that there is likely a 'halo-effect' between the course and instruction ratings. They are both completed by the students in the same survey event. This suggests that effective instructors use effective teaching materials or students' perception of course materials and instruction are not delineated but multiplexed.

The SET outcomes suggest that although the faculty agree or strongly agree that over 90% of the lessons are effective, the qualitative comments imply that the students highlighted several (possible) changes to the lessons. The comments provide more in-depth insights into the lessons and provide useful data to impact institutional policy. Although consultation of the qualitative comments may not directly resolve problems of reliability of using class averages in the SET analysis (Clayson, 2017; Morley, 2012), it provides rich information for policy improvements.

5.2 Faculty Perceptions of the SET

The findings suggest that faculty somewhat agree with the design of the SET and the outcomes for evaluating their teaching. However, faculty, in agreement with Wright (2005) found that the SET insufficiently evaluated (certain') or ('all') aspects of teaching. The findings suggest that faculty perceptions captured by the FSET are not in agreement with the students' ratings of the teaching. The findings thus conflict with Beran and Rokosh 2(009), Idaka et al. (2006) and Inko-Tariah (2013) who suggested that they are similar. This finding negates the dissatisfaction expressed by Iyamu & Aduwa-Oglebaen (2005) and the challenges expressed by Iyam & Aduwa-Oglebaen (2005), Mwachingambi & Wadesango (2011), Boring et al. (2016), Hornstein (2020) and Spooren et al. (2013).

The findings align with Shah et al. (2020) who found that faculty charisma influences the SET scores. Faculty charisma may go a long way in maintaining student attention but does not effectively impact the knowledge gained or the quality of the course materials. Faculty somewhat or strongly agree that the lecture content is more influential in the SET ratings, thus it is a stronger influence on teaching than the faculty charisma as suggested by Shah et al. (2020).

Halo-effect seen in the current SET format and observed by Spooren & Mortelmans (2006) could be removed by separating the course materials and instruction surveys in the SET into independent surveys. Benchmarking of the questions and inclusion of summative and formative assessment questions should be embraced to improve the robustness of the SET. Use of standardized framework for evaluating teaching was somewhat supported by the simplistic and subjective outcomes of the SET. Faculty support for the use of FSET, ICE, and peer evaluations is indicated as hinging on the adoption of a teaching framework. The faculty ranked the SET higher than ICE, FSET

and peer evaluations for the purpose of faculty evaluation. The faculty also suggested that peer evaluations are negatively impacted by internal conflicts, peer embellishments and other challenges.

The findings indicate that the SET is not challenged by student's knowledge of a good lesson. The graduate students of the HEI understand what a "good lesson" is from inductive and deductive knowledge gained by students over their long-term experiences of teaching in higher education. This contrasts with findings by Mohammad (2001) that students' lack of training in course materials is an invalidating factor. In this case HEI context, it is arguable that as customers, graduate students are well positioned to understand effectiveness.

5.3 Are the FSET and SET Outcomes Related?

Like the findings of Marsh (1987) and Beran & Violato, (2005), FSET and SET outcomes are moderately correlated suggesting that the graduate students in this HEI have some understanding of good lessons. Also, the students' perceptions of these lessons moderately align with that of the faculty. However, the robustness of the SET and FSET could be improved by broadening the questions designed to evaluate teaching effectiveness beyond the five questions currently used by the case HEI.

5.4 Justifications for an Inclusive Evaluative Model

Findings in Table 8 also indicate that faculty are evenly split on their favorability of the use of the SET as an evaluative tool. The mean of 3.91 suggests faculty believe that there are other important aspects of teaching skills which could not be evaluated by Likert scale ratings. Like Wright (2005), the research finds that the SET may be inadequate for evaluation of teaching. Inclusion of a qualitative survey section in the SET allows students to reflect on other aspects of teaching not evaluated by the closed ended questions. This allows for student evaluation of the faculty unguided by the HEI's views.

Outcomes seen in figure 1 show that the faculty preference for the SET challenges the view of Molesworth et al. (2009) that students as customers negatively impacts the ability of the HEI to achieve regulatory standards. The point of the faculty is that a more inclusive evaluative approach with inputs from the faculty, students, independent experts and peers or a combination of a few of these should be adopted. In particular, the SET, peer input and FSET outcomes are strongly rated by the faculty. The outcomes of faculty preference of these tools in comparison with the Set particularly shows that the SET is preferred above each of the other tested evaluative tools.

6. Conclusion

The research has produced further evidence in support of the use of SETs in the evaluation of teaching over the use of other tools. However, the faculty support the use of the more inclusive cluster of tools for the evaluation of instructional practices. The weighting of the tools within the model is dependent on the HEI's focus. In other words, the HEI could assign weights to the SET, FSET, ICI and peer review based on faculty ratings of these tools. Although each of these tools have challenges that could impair their outcomes, use of a combination of these tools could combine student assessments, faculty self-evolution, peer evaluation and independent consultant evaluations. A model that inculcates all four tools or a combination of some of these tools would be more administratively demanding but would distribute the weighting away from a simple student input, self-evaluation, peer evaluation, or independent consultant and use of standardized rubrics. This approach could employ input from students, and a combination of faculty or independent consultants. This approach to faculty evaluation is more robust empowering the faculty, students, independent consultants and peers.

The faculty views of the implementation of surveys for evaluating the instruction, and the course materials reveal that the students commonly assign the same ratings across board for each faculty. It thus suggests that the students make up their minds on a score and assign this score across all survey questions. It also shows that the students did not rate any faculty instruction or course materials below average. These scores leave very little room for inferences that could impact policy positions.

7. Implications for HEIs

Although indicative and not conclusive, the findings suggest that integration of FET and SET could be used to enhance the evaluation of teaching quality, curriculum delivery and the evaluation of faculty and student acceptance of the quality of the curriculum. However, the FSET and SET are challenged by questions in the FSET and SET themselves. The implication of the findings is HEIs could include questions with more depth than currently adopted. The FSET and SET are likely challenged by biases in these tools and could produce outcomes misaligned with actual teaching effectiveness. Outcomes of the SET as seen, somewhat correlate with the outcomes of the FSET, however their relationship could be made clearer by improving the test questions.

The faculty suggestion of combining FSET, SET, ICE and peer evaluation tools for the evaluation of faculty would be a more holistic and robust approach than the use of only the SET, FSET or a combination of both. This approach would include student, faculty, within faculty and expert independent evaluations in the review of curriculum, teaching and learning. By combining the FSET, SET, ICE and peer evaluations, the negative effects of each tool could be damped by their advantages. Such approach would include the views of the students, the faculty, peers and independent consultants. It would inculcate the benefits of students and faculty perceptions and standardized teaching framework.

8. Limitations

Due to the sample size of the faculty, the findings are exploratory and non-generalizable beyond the specific case HEI. Also, multi-institutional follow-up study could be done to test the robustness of the findings.

Although faculty preference for the multi-tools approach investigated in this study, future research could investigate the weighting systems, timelines, and integration into faculty promotion mechanisms.

References

Alsmadi, A. (2005). Assessing the quality of students' ratings of faculty members at Mu'tah University. *Social Behavior and Personality, 33*(2), 183-8. https://doi.org/10.2224/sbp.2005.33.2.183

Arbula Blecich, A., & Zaninović, V. (2019). Insight into students' perception of teaching. *Management, 24*(1), 137-152. https://doi.org/10.30924/mjcmi.24.1.9

Arnold, I. J. M. (2009). Do examinations influence student evaluations? *International Journal of Educational Research, 48*(4), 215-224. https://doi.org/10.1016/j.ijer.2009.10.001

Atwater, L., Waldman, D., & Brett, J. (2002). Understanding and Optimizing Multisource Feedback. *Human Resource Management, 41*(2), 193-208. https://doi.org/10.1002/hrm.10031

Bailey, K. D. (1994). *Typologies and taxonomies: an introduction to classification techniques.* Sage. https://doi.org/10.4135/9781412986397

Balam, E., & Shannon, D. (2010). Student ratings of college teaching: A comparison of faculty and their students. *Assessment and Evaluation in Higher Education, 35*, 209-221. https://doi.org/10.1080/02602930902795901

Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of educational psychology, 79*(3), 308. https://doi.org/10.1037//0022-0663.79.3.308

Benton, S. L., & Cashin, W. E. (2012). *Student ratings of teaching: a summary of research and literature* (IDEA Paper no. 50). Manhattan, KS: The IDEA Center. https://ideacontent.blob.core.windows.net/content/sites/2/2020/01/PaperIDEA_50.pdf.

Benton, S.L., & Young, S. (2018). *Best Practices in the Evaluation of Teaching*. IDEA Paper #69.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education, 17*(1), 48-62.

Berk, R. A. (2018). *Peer Review of Teaching: A Sourcebook*. Stylus Publishing.

Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment and Evaluation in Higher Education, 30*, 593-601. https://doi.org/10.1080/02602930500260688

Beran, T. N., & Rokosh, J. L. (2009). Instructor's perspectives on the utility of student ratings of instruction. *Instructional Science, 37*, 171-184. https://doi.org/10.1007/s11251-007-9045-2

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International journal of teaching and learning in higher education, 17*(1), 48-62.

Borgen, J. A., & Davis, D. E. (1974). *Planning, Implementing, and Evaluating Career Preparation Programs*.

Boring, A., Ottoboni, K., & Stark, P.B. (2016). "*Student evaluations of teaching (mostly) do not measure teaching effectiveness*", ScienceOpen Research. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Cashin, W. E. (1990). *Student ratings of teaching: Recommendations for use* (Rep. No. IDEA Paper No. 22). Kansas State University: Center for Faculty Evaluation & Development.

Cashin, W. E. (1995). *Student Ratings of Teaching: The Research Revisited*. IDEA Paper No. 32.

Cashin, W. E. (2003). Evaluating college and university teaching: Reflections of a practitioner. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 531-593). Dordrecht, The Netherlands: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-010-0137-3_10

Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education, 71*, 17-33. https://doi.org/10.1080/00221546.2000.11780814

Charteris, J., & Smardon, D. (2019). The politics of student voice: Unravelling the multiple discourses articulated in schools. *Cambridge Journal of Education, 49*(1), 93-110. https://doi.org/10.1080/0305764X.2018.1444144

Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & evaluation in higher education, 28*(1), 71-88. https://doi.org/10.1080/02602930301683

Chikazinga, W. W. N. (2018). Perceptions of lecturers of student evaluations of their teaching. *International Education Journal, Comparative Perspectives, 17*(4), 36-48.

Chism, N. V. N. (1999). *Peer Review of Teaching. A Sourcebook*. Anker Publishing Company, Inc., 176 Ballville Road, PO Box 249, Bolton, MA 01740-0249.

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review. *Journal of Marketing Education, 31*(1), 16-30. https://doi.org/10.1177/0273475308324086

Clayson, D. E. (2017). Student evaluation of teaching and matters of reliability. *Assessment & Evaluation in Higher Education, 43*(4), 666-681. https://doi.org/10.1080/02602938.2017.1393495

Debroy, A., Ingole, A., & Mudey, A. (2019). Teachers' perceptions on student evaluation of teaching as a tool for faculty development and quality assurance in medical education. *Journal of Education and Health Promotion, 8*. https://doi.org/10.4103/jehp.jehp_47_19

Denson, N., T. Loveday, & H. Dalton. (2010). "Student Evaluation of Courses: What Predicts Satisfaction?" *Higher Education Research & Development, 29*(4), 339-356. https://doi.org/10.1080/07294360903394466

Dolnicar, S., & Grün, B. (2009). Response Style Contamination of Student Evaluation Data. *Journal of Marketing Education, 31*(2), 160-172. https://doi.org/10.1177/0273475309335267

Douglas, J., & Douglas, A. (2006). Evaluating Teaching Quality. *Quality in Higher Education, 12*(1), 3-13. https://doi.org/10.1080/13538320600685024

Entrekin, L., & Chung, Y. (2001). Attitudes towards different sources of executive appraisal: a comparison of Hong Kong Chinese and American managers in Hong Kong. *International Journal of Human Resource Management, 12*(6), 965-987. https://doi.org/10.1080/713769693

Esarey, J., & Valdes, N. (2020). Unbiased, reliable, and valid student evaluations of teaching. *Assessment & Evaluation in Higher Education, 45*(1), 71-84. https://doi.org/10.1080/02602938.2020.1724875

Flick, U. (2009). *An introduction to qualitative research*. Sage.

Ginns, P., Prosser, M., & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: the perspective of currently enrolled students. *Studies in Higher Education, 32*(5), 603-615. https://doi.org/10.1080/03075070701573773

Gustad, J. W. (1961). Policies and practices in faculty evaluation. *Educational Record, 42*, 194-211.

Harun, S., Dazz, S. K., Saaludin, N., & Che Wan Ahmad, W. S. (2011). Lecturers' perception on student Evaluation at Universiti Kuala Lumpur. *Enhancing Learning: Teaching & Learning Conference*, 1-10. https://doi.org/10.1145/2448556.2448611

Henderson, C., Turpen, C., Dancy, M., & Chapman, T. (2014). Assessment of teaching effectiveness: Lack of alignment between instructors, institutions, and research recommendations. *Physical Review Special Topics-Physics Education Research, 10*(1), 010106. https://doi.org/10.1103/PhysRevSTPER.10.010106

Hou, A. Y. C., Hill, C., Lin, A. F. Y., & Justiniano, D. (2022). The battle for legitimacy of 'student involvement'in external quality assurance of higher education from Asian QA perspectives–does policy contextualization matter? *Journal of Asian Public Policy*, 1-19. https://doi.org/10.1080/17516234.2022.2092385

Hudson, W. W. (1982). *The clinical measurement package: A field manual.* Homewood, IL: Dorsey.

Idaka, I. I., Joshua, M. T., & Kritsonis, W. A. (2006). Attitude of academic staff in Nigerian tertiary educational institutions to student evaluation of instruction (SEI). *National Forum of Educational Administration and Supervision Journal, 23*(4).

Inko-Tariah, D. C. (2013). Attitude of lecturers towards student evaluation of their teaching effectiveness in Nigeria universities. *Journal of Education and Practice, 4*(15).

Iyam, E. O. S., & Aduwa-Oglebaen, S. E. (2005). Lecturers perception of student evaluation in Nigerian universities. *International Educational Journal, 6*(5), 619-625.

Johnson, R. (2000). The authority of the student evaluation questionnaire. *Teaching in Higher Education, 5*, 419-434. https://doi.org/10.1080/713699176

Keeley, R. G. (2015). Measurements of student and teacher perceptions of co-teaching models. *The Journal of Special Education Apprenticeship, 4*(1), 4. https://doi.org/10.58729/2167-3454.1042

Kember, D., Leung, D., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment and Evaluation in Higher Education, 27*, 411-425. https://doi.org/10.1080/0260293022000009294

Kulik, J.A. (2001). "Student Ratings: Validity, Utility, and Controversy." *New Directions for Institutional Research, 2001*(109), 9-25. https://doi.org/10.1002/ir.1

Lattuca, L. R., & Stark, J. S. (2009). *Shaping the college curriculum: Academic plans in context*. John Wiley & Sons.

Lyde, A. R., Grieshaber, D. C., & Byrns, G. (2016). Faculty teaching performance: Perceptions of a multi-source method for evaluation. *Journal of the Scholarship of Teaching and Learning, 16*(3), 82-94. https://doi.org/10.14434/josotl.v16i3.18145

Marsh, H. W. (1987). Student's evaluations of university teaching: Research findings, methodological issues, and directions for further research. *International Journal of Educational Research, 11*, 253-388. https://doi.org/10.1016/0883-0355(87)90001-2

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52*(11), 1187-1197. https://doi.org/10.1037/0003-066X.52.11.1187

MacBeath, J. (2003). Teacher self-evaluation. *In International handbook of educational research in the Asia-Pacific region: Part one* (pp. 767-780). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-3368-7_53

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52*, 1218-1225. https://doi.org/10.1037/0003-066X.52.11.1218

Molesworth, M., Nixon, E., & Scullion, R. (2009). Having, being and higher education: The marketisation of the university and the transformation of the student into consumer. *Teaching in higher Education, 14*(3), 277-287. https://doi.org/10.1080/13562510902898841

Moreno, M. L., Santos, N. R. D., & Pais, L. M. (2021). Empirical research on multisource feedback in performance appraisal: a literature review. *Revista Psicologia Organizações e Trabalho, 21*(4), 1739-1749. https://doi.org/10.5935/rpot/2021.4.21847

Morley, D. D. (2012). "Claims about the Reliability of Student Evaluations of Instruction: The Ecological Fallacy Rides Again." *Studies in Educational Evaluation, 38*(1), 15-20. https://doi.org/10.1016/j.stueduc.2012.01.001

Mwachingambi, S., & Wadesango, N. (2011). University lecturers' perceptions of student evaluation of their instructional practices. *Anthropologist, 13*(3), 167-174. https://doi.org/10.1080/09720073.2011.11891194

Onwuegbuzie, A. J., Daniel, L. G., & Collins, K. M. T. (2009). A meta-validation model for assessing the score validity of student teaching evaluations. *Quality & Quantity, 43*(2), 197-209. https://doi.org/10.1007/s11135-007-9112-4

Park, K. A., Bavis, J. P., & Nu, A. G. (2020). *Branching Paths: A Novel Teacher Evaluation Model for Faculty Development*.

Penny, A. R. (2003). Changing the Agenda for Research into Students' Views about University Teaching: Four shortcomings of SRT research. *Teaching in Higher Education, 8*(3), 399-411. https://doi.org/10.1080/13562510309396

Richardson, J.T.E. (2005). "Instruments for Obtaining Student Feedback: A Review of the Literature." *Assessment & Evaluation in Higher Education, 30*(4), 387-415. https://doi.org/10.1080/02602930500099193

Saldaña, J. (2013). *The Coding Manual for Qualitative Researchers (2nd ed.)*. London: Sage.

Schultz, K. K., & Latif, D. (2006). The planning and implementation of a faculty peer review teaching project. *American Journal of Pharmaceutical Education, 70*(2). https://doi.org/10.5688/aj700232

Shah, S., Kumar, A., & Pokharel, N. (2020). Evaluation of Overall Performance of Faculty Members by Using Self-Assessment Method. *Journal of Universal College of Medical Sciences, 8*(02), 87-89. https://doi.org/10.3126/jucms.v8i02.34311

Spencer, K. J., & Schmelkin, L. P. (2002). Student perspectives on teaching and its evaluation. *Assessment and Evaluation in Higher Education, 27*, 397-409. https://doi.org/10.1080/0260293022000009285

Spooren, P. (2010). On the credibility of the judge: A cross-classified multilevel analysis on students' evaluation of teaching. *Studies in educational evaluation, 36*(4), 121-131. https://doi.org/10.1016/j.stueduc.2011.02.001

Spooren, P., Brockx, B., & Mortelmans, D., (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*, 598-642. https://doi.org/10.3102/0034654313496870

Spooren, P., & Mortelmans, D. (2006). Teacher professionalism and student evaluation of teaching: Will better teachers receive higher ratings and will better students give higher ratings? *Educational Studies, 32*, 201-214. https://doi.org/10.1080/03055690600631101

Sulong, M. S. (2014). Lecturers' perception on teaching evaluation: Selection of research instruments. *The Online Journal of Quality in Higher Education, 1*(2), 36-44.

Theall, M., & Keller, J.M. (2017). Summary and Recommendations. *Teaching and Learning, 2017*, 109-112. https://doi.org/10.1002/tl.20273

ÜSTÜNLÜOĞLU, E., & Seda, C. A. N. (2012). Student evaluation of teachers: A case study at tertiary level. *International journal on new trends in education and their implications, 3*(4). https://doi.org/10.5296/ije.v4i3.1811

Vogt, W. Paul. (1999). *Dictionary of statistics and methodology*. Sage: Thousand Oaks, California.

Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education, 23*, 191-212. https://doi.org/10.1080/0260293980230207

Wright, M. (2005). Always at odds?: Congruence in faculty beliefs about teaching at a research university. *The Journal of Higher Education, 76*(3), 331-353. https://doi.org/10.1080/00221546.2005.11772285

Yin, H., W. Wang, & J. Han. (2016). "Chinese Undergraduates' Perceptions of Teaching Quality and the Effects on Approaches to Studying and Course Satisfaction." *Higher Education,* 71(1), 39-57. https://doi.org/10.1007/s10734-015-9887-5